US005710549A

United States Patent [19]

Horst et al.

[11] Patent Number: 5,710,549

[45] Date of Patent: Jan. 20, 1998

[54] **ROUTING ARBITRATION FOR SHARED RESOURCES**

[75] Inventors: **Robert W. Horst**, Saratoga, Calif.; **William J. Watson**; **David P. Sonnier**, both of Austin, Tex.

[73] Assignee: **Tandem Computers Incorporated**, Cupertino, Calif.

[21] Appl. No.: **483,663**

[22] Filed: **Jun. 7, 1995**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 316,431, Sep. 30, 1994, abandoned.

[51] **Int. Cl.**[6] **H04Q 1/18**

[52] **U.S. Cl.** **340/825.5**; 370/462

[58] **Field of Search** 340/825.5, 825.51; 370/444, 455, 462, 461, 447; 395/291, 296, 303, 650, 728, 729, 731, 732, 299, 860, 861, 862

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,699,524	10/1972	Norberg	340/825.5
4,663,756	5/1987	Retterath	340/825.5
5,072,363	12/1991	Gallagher	395/725
5,210,871	5/1993	Lala et al.	395/650
5,388,097	2/1995	Baugher et al.	370/455
5,388,228	2/1995	Heath et al.	395/303
5,392,033	2/1995	Oman et al.	340/825.5
5,479,158	12/1995	Sato	340/825.5

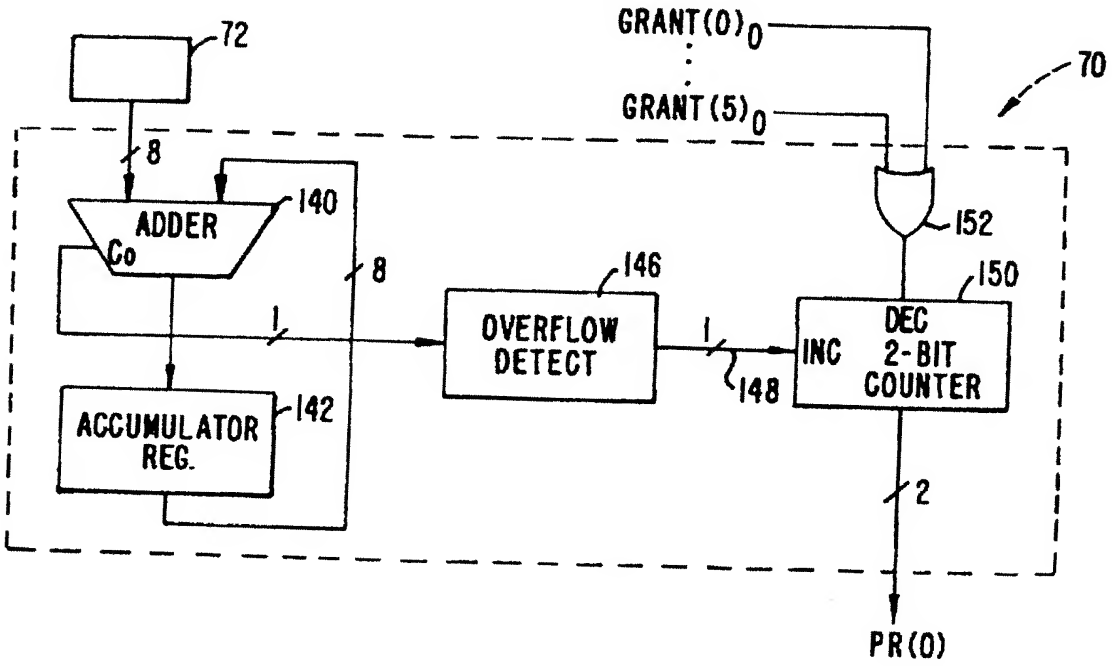
Primary Examiner—Edwin C. Holloway, III

Attorney, Agent, or Firm—Townsend and Townsend and Crew LLP

[57] **ABSTRACT**

A data communicating device, having a number of inputs whereat data is received for communication from one of a number of outputs of the device, includes apparatus for providing two levels of arbitration to select one of the inputs for data communication to an output. The first (lower) level of arbitration bases selection upon a round-robin order; the second (higher) arbitration level selects inputs based upon an indication from an input of an undue wait for access to the output over a period of time. Each input is provided a modulo-N counter, and a digital counter. Each time an input contends for access to an output and loses to selection by the output to another input, the modulo-N counter is incremented by an assigned value for that input. When N is exceed without access, the digital counter is incremented. The content of the counter operates to force the high-level arbitration.

13 Claims, 3 Drawing Sheets



(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号

特開平9－138774

(43)公開日　平成９年(1997)５月27日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ	13/362	５１０		Ｇ０６Ｆ	13/362	５１０Ｇ
	7/50				7/50	Ｋ
	15/173				15/16	４００Ｔ

審査請求　未請求　請求項の数４　ＯＬ　（全 10 頁）

(21)出願番号　特願平8－145734

(22)出願日　平成８年(1996)６月７日

(31)優先権主張番号　０８／４８３６６３

(32)優先日　1995年６月７日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　391058071
タンデム　コンピューターズ　インコーポレイテッド
TANDEM COMPUTERS INCORPORATED
アメリカ合衆国　カリフォルニア州　95014　クーパーティノ　ノース　タンタウ　アベニュー　10435

(72)発明者　ロバート　ダブリュー　ホースト
アメリカ合衆国　カリフォルニア州　95070　サラトガ　ラーチモント　アベニュー　12386

(74)代理人　弁理士　中村　稔　（外６名）

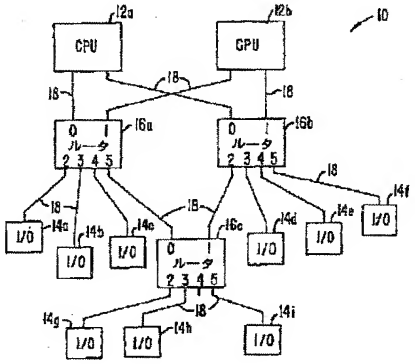
最終頁に続く

(54)【発明の名称】　共用リソースのためのルート裁定方法

(57)【要約】

【課題】　同じ出力へのデータ通信アクセスについて競合する２つ以上の入力の裁定にバイアスを与える方法及び装置を提供する。

【解決手段】　多数の出力の１つから通信するためにデータ受け取られる多数の入力を有したデータ通信装置は、出力へデータを通信するための入力の１つを選択するように２レベルの裁定を行う装置を備えている。第１の（低い）レベルの裁定は、ラウンドロビン順序に基づいて選択を行い、第２の（高い）裁定レベルは、ある時間周期にわたり出力へのアクセスが不当に待機された入力からの指示に基づいて入力を選択する。各入力には、モジュロＮカウンタ及びデジタルカウンタが与えられる。入力が出力へのアクセスに競合し、それに敗れて、出力が別の入力を選択するたびに、モジュロＮカウンタがその入力の指定値だけ増加される。アクセスせずにＮを越えると、デジタルカウンタが増加される。カウンタの内容は、高レベル裁定を強制するように動作する。



1

【特許請求の範囲】

【請求項１】　少なくとも一対の入力と、出力とを有し、その一対の入力にメッセージデータを受け取って、出力へ通信しそして出力から再送信するように動作するデータ通信装置において、出力へのアクセスに競合する一対の入力間の裁定にバイアスを与える方法が、

各々の入力に指定値を与え、

一対の入力各々の値が等しいときは第１の所定のベースで出力へ通信するように一対の入力の一方を選択し、

一対の入力各々の値が等しくないときは第２の所定のベースで出力へ通信するように一対の入力の一方を選択し、そして出力へ通信するためのメッセージデータを有する入力の一方の値を変更する、という段階を備えたことを特徴とする方法。

【請求項２】　上記変更段階は、上記指定値により変更される各入力の累積値を発生する請求項１に記載の方法。

【請求項３】　複数の入力と、少なくとも１つの出力とを有し、複数の入力にデータを受け取って出力から再送信するように動作するデータ通信装置であって、出力へそしてそこから通信するためのデータを有する多数の複数の入力の中から選択を行うデータ通信装置において、複数の入力の各々に対し、

（ａ）指定値を受け取りそしてそこから変更された値を発生するように接続された演算ユニットを備え、この変更された値は、上記複数の入力の１つが出力のためのデータを有しそして上記複数の入力の別のものが選択されたときに上記指定値によって変更され、

（ｂ）上記演算ユニットに接続され、上記変更された値が所定値に等しいか又はそれを越えるときにカウントを増加するためのカウンタを備え、

上記出力は、上記複数の入力の各々からカウントを受け取って、そのカウントが第１の値であるときは第１の順序に基づき多数の入力の１つから出力へデータを通信するように多数の入力の１つを選択し、そして複数の入力のいずれかからのカウントが第１の値でないときは第２の順序に基づき多数の入力の１つを選択するためのアービタロジックを有することを特徴とするデータ通信装置。

【請求項４】　上記演算ユニットは、その演算ユニットにより発生された和が桁上げを生じるときに桁上げ信号がアサートされる桁上げ出力を含む請求項６に記載の装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、一般に、多数の入力と出力を有していて通信ネットワーク内でメッセージのトラフィックをルート選択する装置に係る。より詳細には、本発明は、装置の同じ出力へのデータ通信アクセスに対して競合する装置の２つ以上の入力間の裁定にバイアスを与える方法及びこのような方法を実施する装置に係る。

2

【０００２】

【従来の技術】マルチ処理環境が増大するにつれて、システムの個々のシステム要素（即ち、プロセッサや周辺装置）間でデータ通信を行う機能としては、通信経路又は他の幾つかの共用リソース（例えば、マルチユーザバスシステム）への不公平なアクセスをあるシステム要素に偶発的に与えないようにすると同時に、他のシステム要素へのこのようなアクセスを制限するように入念に考えねばならない。今日の多数のデータ通信ネットワーク構成は、１つの通信リンクから多数の他の通信リンクの１つへメッセージトラフィックを向け又はルート指定するために種々の装置（典型的に「ルータ」）を使用する。しかしながら、メッセージトラフィックは、同じ出力に向けて実質的に同時に装置の２つ以上の入力に受け取ることができ、２つのうちのどれを他の前に処理できるようにするか、即ちどれを最初に行かせるかの闘争が生じる。出力へのアクセスは、２つの競合する入力間にある形態の裁定を必要とする。小型のシステム（即ち、互いに通信するために少数のシステム要素しか必要としないシステム）は、入力に固定の優先順位を指定したり又は「ラウンドロビン」処理を使用したりといった古典的な裁定技術を用いることができる。前者の場合には、各ポート入力に、ハイアラーキ裁定機構を形成するための所定優先順位が与えられる。高い優先順位の入力は、その同じ出力へアクセスを求める入力であって低い優先順位が予め指定された入力以上に出力へのアクセスが与えられる。ラウンドロビン技術は、入力の間に所定の順序に基づいて優先順位を指定することを含む。メッセージトラフィックを受け取ると、優先順位が変化し、特定の出力へのアクセスが許可された最後の入力に最低の優先順位が与えられ、その次の順の入力が最高の優先順位を有し、残りの入力は、所定の順序に基づいて同様に変化する優先順位を有する。

【０００３】例えば、マルチユーザＶＭＥバスシステムのような他の共用リソース環境に同様の技術を用いて、バスへのアクセスを、接続されたコントローラ間に割り当てることもできる。

【０００４】

【発明が解決しようとする課題】コンピュータシステムが複雑になり、要素（プロセッサ及び周辺ユニット）の数が増加するにつれて、ルート指定装置の入力が多数の要素からのメッセージトラフィックをルート指定するよう要求されることが例外ではなくなった。更に、ルート指定装置は、例えば、ツリー状のネットワーク構成とするように時々カスケード接続にされ、ルート指定装置の入力を経てメッセージトラフィックをルート指定する必要のあるシステム要素の数が増加する。予め指定された優先順位、ラウンドロビン処理、又は他の同様の技術を

使用する場合には、ルート指定装置のある入力のサービスがシステム要素の幾つかに不公平に割り当てられ、そのルート指定装置の別の入力を使用する他の要素を犠牲にすることがある。例えば、あるルート指定装置は、１つのシステム要素のみからメッセージトラフィックを１つの入力に受け取るが、同じルート指定装置の別の入力は、多数の要素にサービスすることが要求される。上記技術のいずれかを使用すると、ルート装置の注意の多くが１つの要素に向けられ、第２の入力を用いる多数の要素の各々にはあまり注意が向けられない。従って、公知の裁定技術は、ルート指定装置のサービスの多くを、そのルート指定装置を使用する少数のシステム要素を有する入力に不公平に割り当てることになる。

【０００５】

【課題を解決するための手段】本発明は、例えば、共用バス構造体のような任意の形式の共用リソース、又はここに述べるネットワークルート指定装置の出力へのアクセスを求める多数のユーザ間を裁定する方法を提供する。本発明は、多数のメッセージ受信ポート入力と、少なくとも１つの出力ポートとを有し、上記ポート入力の１つにメッセージトラフィックが受け取られて、その出力ポートへルート指定されそしてそこから送信されるようなネットワークルート指定装置に使用される好ましい実施形態について説明する。本発明は、各ポート入力を通るメッセージトラフィックを監視して、待機中のメッセージトラフィックを有するポート入力の優先順位を上昇するという考え方に基づいている。本発明の好ましい実施形態によれば、２レベルの裁定が実施される。最初に、低レベル優先順位機構を用いて、ポート入力へのアクセスを求める２つのポート入力間で裁定が行われ、多数の裁定周期を通じて待機していたメッセージトラフィックを有するポート入力を受け入れるために、それらの優先順位が上昇され、高い優先順位機構を瞬間的に使用する裁定へと持っていかれる。低レベル機構は簡単なラウンドロビン優先順位を使用する。ポート出力へのアクセスをもつ最後のポート入力は、裁定の優先順位を有し、一方、次の順番のポート入力が最高の優先順位を有する。他のポート入力は、確立されたラウンドロビン順序に基づいて下降する優先順位を有する。ポート入力がポート出力へのアクセスを得るたびに優先順位が変化する。

【０００６】第１レベルにおいて裁定を行うときには、各々の選択されたポート入力は、その受け取ったメッセージトラフィックを送信のために出力へルート指定し、優先順位がラウンドロビン機構に基づいて進んで、そのポート入力の優先順位を最低にし、そしてラウンドロビンシーケンスで次のポートを最高の優先順位をもつものとして指名する。高レベルの裁定機構は、ポート出力を通るメッセージトラフィックを監視しそしてポート出力へルート指定されるべく待機しているメッセージトラフィックを有するポート入力の優先順位を高める。高められる量は、各ポート入力に対して予め指定されたバイアス値に一部依存する。予め指定されたバイアス値は、入力分数（ＩＦ）の形態であり、これは、ポート出力の帯域巾の一部分のポート入力の割り当てとして考えられる。ポート出力により取り扱われるメッセージトラフィックの流れが監視され、待機中のメッセージトラフィックを有するポート入力の優先順位は、ポート入力が待機しているポート出力へのアクセス及び裁定に勝つことができないたびに高めることができる。そのポート入力に指定されたＩＦにより時間間隔が確立される。ポート入力が出力ポートの裁定に参加する各裁定周期は、予め指定された値でモジュロＮカウンタを増加する。このカウンタがオーバーフローすると、２ビットカウンタが増加される。この２ビットカウンタの非ゼロカウントは、関連ポート入力がポートへ優先順位要求を発行し、その優先順位が高められ、そして順番が変えられる(out of turn) という信号である。２つ以上のポート入力がその関連２ビットカウンタに非ゼロカウントを有する場合には、高レベル機構が、大きい方の非ゼロカウントをもつものへアクセスを許可する。２つ以上のポート入力の２ビットカウンタが非ゼロであって且つ等しい場合には、固定の優先順位機構で裁定が行われる。２ビットカウンタに非ゼロカウントを有するポート入力がアクセスを許可されるたびに、２ビットカウンタが減少される。

【０００７】従って、優先順位カウンタに０以外のカウントを有するポート入力を見る裁定周期が最初に裁定され、同点の場合は、固定優先順位によって行われる。各ポート入力のＩＦ値を形成する指定のバイアス値は、モジュロＮカウンタ構成を増大するのに使用される。（ここで、Ｎは２５６であり、従って、カウンタは２５５でロールオーバーする。）カウンタがその最大カウント（２５５）を越えて増加されそしてロールオーバーすると、２ビットカウンタが１だけ増加される。本発明により多数の効果が達成される。第１に、ルート指定装置のポート出力への公平なアクセスが、いかなるポート入力にも与えられ、即ちシステム要素に直結されたポート入力及びシステム要素に間接的に接続された（即ち他のポートを経て）ポート入力の両方に与えられる。第２に、本発明は、ポート出力の最大帯域巾の最小保証部分をいかなるポート入力にも割り当てる。他のメッセージトラフィックを待機してデータが失われないよう確保するために、例えば、リアルタイムデータ（例えば、ビデオ）を取り扱うポート入力にポート出力へのアクセスを傾斜させるように、高い指定バイアス値を与えることができる。第３に、全ネットワークについて指定のバイアス値を与えることにより、いかなる２つのシステム要素間にも保証されたメッセージ送信待ち時間を確立することができる。これは、ネットワークの混雑によるのではないエラー状態のもとでのみ越える特定の値に低い時間切れ

値をセットできるようにする。

【０００８】本発明のこれら及び他の効果は、添付図面を参照した本発明の以下の詳細な説明より当業者に容易に明らかとなろう。

【０００９】

【発明の実施の形態】添付図面の図１には、簡単なマルチプロセスシステムが参照番号１０で一般的に示されている。図示されたように、このマルチプロセスシステム１０は、少なくとも一対の中央処理ユニット（ＣＰＵ）１２ａ、１２ｂと、ルート指定ユニット即ちルータ１６及び両方向性通信リンク１８によりシステムエリアネットワーク構成に相互接続された複数の入力／出力ユニット１４（１４ａ、１４ｂ、・・・１４ｉ）とを備えている。システム１０の種々の要素間のメッセージトラフィックは、直列送信される９ビット記号と、これら記号の同期転送のために必要な送信クロックとを含むデータパケットの形態であるのが好ましい。これらの記号は、ネットワークプロトコルの流れ制御に使用されるデータ又はコマンドを形成するようにエンコードされる。ネットワーク流れ制御は、本発明の理解又は実施に関与せず、従って、ルート指定ユニット１６の幾つかの要素の説明に必要な以外は、ここでは詳細に述べない。しかしながら、各メッセージは、メッセージのソース及び行先を識別するデータを含む。行先は、メッセージが再送信されてくるところのポート出力を選択するためにルータ１６により使用される。

【００１０】図１の説明を続けると、各々のルータ１６は、６個の両方向ポート（０、１、・・・５）を有し、その各々は、メッセージトラフィックが受け取られるポート入力と、メッセージトラフィックを送り出すことのできるポート出力とを有している。ルータ１６ａのポート２、３、４（及び０、１）の各々は、１つのシステム要素にのみサービスする（即ち、その要素からのトラフィックをルート指定する）。一方、ポート５は、このポートを経てルート指定されるメッセージトラフィックを送信できる８つのシステム要素、即ち両ＣＰＵ１２（ルータ１６ｂ及び１６ｃを経て）と、６つのＩ／Ｏユニット１４（ルータ１６ｂ、１６ｃを経て）とを有する。これら８個全てのソースが、ルータ１６ａのポート５のポート入力を経てルート指定されるべきメッセージトラフィックを送信し、ルータ１６ａのポート出力、例えば、ポート０のポート出力（０）を経て送信しようとすることがある。これに対し、ルータ１６ａのポート１－４は、ポート０のポート出力へのアクセスに競合する必要のある単一の要素しか有していない。ポート出力への等しいアクセスが各ポート入力に許可される裁定方法では、ポート２、３及び４に各々接続されたＩ／Ｏユニット１４ａ、１４ｂ及び１４ｃ各々の方が、ルータ１６ａのポート５にメッセージトラフィックを送信するＩ／Ｏユニット１４ｇ、１４ｈ及び１４ｉよりもポート０へのアクセスがより頻繁に与えられる。本発明は、ポート２－４よりも頻繁にポート０（又は他のポート）へのアクセスを与えるようにルータ１６ａのポート５をバイアスすることによりこの問題を軽減するように作用する。

【００１１】図２はルータ１６ａの簡単なブロック図である。ルータ１６ｂ及び１６ｃは、特に指示のない限り、ルータ１６ａと実質的に同じに構成され、従って、ルータ１６ａについての以下の説明がルータ１６ｂ、１６ｃにも等しく適用されることが明らかである。上記のように、ポート０、１、・・・５の各々は、メッセージトラフィックを送信及び／又は受信することができる。それ故、図２は、ルータ１６ａが、各ポート０、１、・・・５に対し、メッセージトラフィックを受信するためのポート入力（Ｉ）と、メッセージトラフィックを送出するためのポート出力（Ｏ）とを有するものとして示している。各ポート入力は、メッセージトラフィックの受信を処理するための関連入力ロジック３０（30_0、30_1、・・・30_5）と、メッセージトラフィックを送出する出力ロジック３２（32_0、32_1、・・・32_5）とを有する。到来するメッセージトラフィックは、受信ポートの入力ロジック３０から、クロスバースイッチ３４によりポート出力の１つへルート指定され、クロスバースイッチは、制御・状態ロジック３６（及び以下に述べる個々の出力ロジック要素３２）によって一部制御される。従って、例えば、ポート０のポート入力（Ｉ）により受け取られるメッセージトラフィックは、それに関連する入力ロジック３０。へ送られ、そしてクロスバースイッチ３４により指定の出力ロジック（例えば、出力ロジック30_3）へルート指定される。ポート３のポート出力Ｏ（３）は、これに接続されたデータを送信するための出力ロジック32_3を有する。

【００１２】制御・状態ロジック要素は、ルータのほとんどの動作に対して同期制御を与える種々の状態マシンを含んでいる。更に、ルータ１６ａは、ルータの要素を同期動作するのに必要な種々のクロック信号を供給するクロックロジック４０と、１つの例外を除いてここでは本発明に関連しない幾つかの自己チェック動作を実行するための自己チェック回路４２とを備えている。ルータ１６ａには、これをメンテナンス処理システム（図示せず）へ通信接続するためにオンラインアクセスポート（ＯＬＡＰ）４６が設けられる。このＯＬＡＰ４６は、以下に述べるように、ルータが、例えば、各ポート入力の指定のバイアス値のような種々の動作情報を受信できるようにするインターフェイスをメンテナンス処理システムに与える。ＯＬＡＰ４６は、ＩＥＥＥ規格１１４９．１に合致するよう構成されたシリアルバス４８に接続される。従って、始動時又は動作進行中に情報がルータ１６ａに与えられる。当業者に明らかなように、ＩＥＥＥ規格１１４９．１は、ＩＥＥＥ規格１１４９．１－１９９０年版、１９９０年５月２１日、ＳＨ１３１４

４、インスティテュート・オブ・エレクトリカル・アンド・エレクトロニック・エンジニア、３４５イースト、４７番ストリート、ニューヨーク、ニューヨーク州、１００１７をベースとするものである。更に詳細な情報は、この規格を参照されたい。

【００１３】図３は、ポート入力Ｉ（０）の入力ロジック３０。を示すブロック図である。他のポート入力Ｉ（１）、・・・Ｉ（５）の入力ロジック30_1－30_5も実質的に同じ構造であり、特に指示のない限り、入力ロジック３０。の説明を、入力ロジック30_1－30_5の説明と考えられたい。図３に示すように、入力ロジック３０。は入力レジスタ５０を備え、これは、到来するメッセージトラフィックを受信してバッファし、入力の先入れ先出しバッファ待ち行列（ＦＩＦＯ）５２へ転送するように動作する。ＦＩＦＯ５２は、送信エンティティにおいて発信されてレジスタ５０及びＦＩＦＯ５２へのデータをクロックするのに使用されるクロック信号（図示せず）と、ＦＩＦＯ５２から記号を引っ張るのに使用される（ローカル）クロックとの間の同期を与えるように動作する。入力ＦＩＦＯ５２からの情報は、９－８（ビット）コンバータ５４へ送られ、これは、各々の９ビット記号をそのエンコード形態からバイト形態へ変換する。更に、入力ＦＩＦＯ５２からの出力は、コマンドデコード要素５６及びプロトコル・パケットチェックユニット５８へ接続される。コマンドデコードユニット５６は、各記号を検査して、それが流れ制御コマンドであるかどうか、ひいては、ルータが作用を与えねばならないコマンド、又はルータにより作用を受ける必要のあるデータ（適切なポート出力へルート指定するのではない）であるかどうか判断する。プロトコル・パケットチェックユニット５８は、パケットが必要な転送プロトコルを満足するよう確保すると共に、パケットの結果のチェック和をチェックして、パケットがルータ１６ａへ適切に送信されるよう確保するように動作する。もしそうでなければ、プロトコル・パケットチェックユニット５８は、パケットの終わりに、そのパケットをおそらく誤りであると識別する記号を付加する。

【００１４】９－８ビットコンバータ５４を通過した到来メッセージトラフィックは、ＦＩＦＯ制御器６４によって制御される弾力性ＦＩＦＯ６２に受け取られて一時的に記憶される。ＦＩＦＯ６２は、到来メッセージパケットの行先ＩＤを検査できるようにし且つポート出力がクロスバースイッチ３４を操作してメッセージトラフィックをルート指定する時間を許すに充分な一時的記憶容量を備えている。又、ＦＩＦＯ６２は、受信ポート入力が待機しなければならない場合には到来メッセージトラフィックの送信を停止するに充分な時間を許すに足る記憶容量を備えていなければならない。しかしながら、適切なポート出力の選択は、到来メッセージパケットに含まれた行先アドレスによって左右される。この決定は、

到来メッセージパケットに含まれた行先アドレスを受け取るポート出力選択ロジック６６により行われる。この行先アドレスから、ポート出力選択ロジック６６は、指定のポート出力を識別し、６本の要求ラインＲ（Ｏ）ｍの１つにその要求されたポート出力を識別する要求信号をアサートする。但し、ｎ＝０、１、・・・５である。説明を続ける前に、表示法について述べる。上記のように、出力ポート選択ロジック６６は、６本の要求ラインＲ（Ｏ）。、Ｒ（Ｏ）$_1$、・・・Ｒ（Ｏ）$_5$の１つにおいて各々撤送される６個の出力信号を発生する。要求信号ラインの形態は、Ｒ（ｎ）$_m$であり、但し、ｎ（ｎ＝０、１、・・・５）は、信号ラインのドライブソースを識別し、そしてｍ（ｍ＝１、２、・・・５）は、撤送される信号の行先を識別する。従って、ポート出力選択ロジック６６は、６本の要求ラインＲ（Ｏ）。、Ｒ（Ｏ）$_1$、・・・Ｒ（Ｏ）$_5$を駆動し、その各々は、それが撤送する信号を出力ロジック３２。、32_1、・・・32_5へと各々接続する。同様に、各ポート出力の出力ロジック３２は、受信した要求信号に応答して、６本の信号ラインＧＲＡＮＴ（ｎ）$_m$の１つにＧＲＡＮＴ信号をアサートすることによりアクセスを許可する。この場合も、ｎは信号ラインを駆動する出力ロジックを識別し、そしてｍはその駆動信号を受信する入力ロジックを識別する。特に指示のない限り、この説明全体を通じてこの表示法を使用する。

【００１５】図３の説明を続けると、ポート入力Ｉ（０）により受け取られ、例えば、ポート出力Ｏ（３）を識別する行先アドレスをもつ到来メッセージは、要求信号ラインＲ（Ｏ）$_3$に要求をアサートして、ポート出力Ｏ（３）（より正確には、関連出力ロジック32_3）に、該ポート出力に向けられたメッセージトラフィックがポート入力Ｉ（０）に待機していることを知らせる。この要求信号を受け取るポート出力は、次いで、そのアクセスを許可することを表す許可信号を許可信号ラインＧＲＡＮＴ（３）。に発生することで応答する。要求された出力ロジック３２がアクセスを許可すると（以下に詳細に述べる）、クロスバースイッチ３４を通る指定のルートが形成され、メッセージパケットは、弾力性ＦＩＦＯ６２から要求された出力ロジックへルート指定される。又、入力ロジックは、バイアスレジスタ７２の内容を受け取るバイアスロジック７０も備えている。バイアスレジスタ７２は、上記のように、その関連ポート入力Ｉ（０）に対する指定のバイアス値であって、ポート出力の帯域巾についてのそのポート入力の部分を表す指定のバイアス値を受け取る。バイアスレジスタ７２の内容から、バイアスロジック７０は、ポート入力Ｉ（０）（待機中のメッセージトラフィックをもつ）が参加して負け、その優先順位を実際に加速する優先順位要求を発生するところの裁定を監視する。この優先順位要求は、６つ全部のポート出力の入力ロジック３２に接続された

２ビットバスＰＲ（０）により所望のポート出力の出力ロジック３２へ通信される。バイアスロジックは、６つのポート出力の出力ロジック３２から、ＧＲＡＮＴ信号ラインＧＲＡＮＴ（ｎ）$_m$により搬送される許可信号を受け取る。

【００１６】２つ以上のポート入力Ｉ（０）・・・Ｉ（５）が、同じポート出力（例えば、Ｏ（３））を識別する行先アドレスをもつメッセージトラフィックをほぼ同時に受信し始める場合には、どのポート入力を最初に処理しそしてどれを待機させるかについてある決定を行わねばならず、即ち所望のポート出力へのアクセスを裁定して、どのポート入力を最初に行かせそしてどれを待機させるかを決定しなければならない。本発明によれば、裁定は、２つのレベルで行われる。最初に、低レベル裁定が使用され、競合するポート入力が単純なラウンドロビンプロセスによって選択される（が、例えば、固定の優先順位を指定するような他の裁定構成も使用できることが明らかであろう）。多数の裁定及び要求を通じて待機したメッセージトラフィックを有するポート入力が順番を変える優先順位要求を発行することにより高レベル優先順位機構に入る。ラウンドロビン裁定プロセスは、各ポート出力Ｏ（０）、Ｏ（１）、・・・Ｏ（５）により、ポート出力へルート指定されるべき待機中のメッセージトラフィックを有するポート入力から受け取った要求信号Ｒ（ｎ）$_m$に応答して実施される。ポート入力信号がその関連優先順位要求をアサートすることにより順番を変えられるべきであるときは、高レベル裁定機構が強制動作される。明らかなように、ポート入力が、ポート出力へルート指定されるべく待機しているメッセージトラフィックを有するときには、そのポート出力に対して参加する裁定を監視する。待機が続くと、各ポート入力に与えられて入力ロジック３０のバイアスレジスタ７２（図３）に維持された上記の入力分数（ＩＦ）からバイアスロジック７０により優先順位要求が発生される（以下に述べるように）。

【００１７】各ポート入力（Ｉ（０）、Ｉ（１）・・・Ｉ（５））からの２ビット優先順位要求は、優先順位要求バス（PR_0、PR_1・・・PR_5）によってポート出力（Ｏ（０）、Ｏ（１）・・・Ｏ（５））へ接続される。多数のポート入力が１つのポート出力に対して待機中のメッセージトラフィックを有し、それ故、そのポート出力へのアクセスを張り合っており、そしてそれらの各々の優先順位要求がゼロである場合には、裁定が行われ、競合するポート入力の１つがラウンドロビンプロセスを用いて選択される。一方、張り合っているポート入力の１つが非ゼロの優先順位要求を発行する場合には、そのポート入力が高い優先順位を有するものとして処理され、次の裁定周期中にアクセスの順番変更が許可される。２つ以上のポート入力が順番の変更を要求している場合には、固定の優先順位ベースで非ゼロの基準要求を

有するポート入力間でポート出力により裁定が行われる。優先順位要求をいかに発生して使用するかを説明する前に、先ず初めに、ポート出力（Ｏ（０）、Ｏ（１）・・・Ｏ（５））のアーキテクチャーを理解するのが有用であろう。図４は、ポート出力Ｏ（３）の出力ロジック３２$_3$のアーキテクチャーを簡単な形態で示すものである。他のポート出力Ｏ（０）－Ｏ（２）及びＯ（４）－Ｏ（５）に対する出力ロジック３２は、実質的に同じ構造である。図４に示したように、クロスバースイッチ３４の出力は、出力ロジック３２$_3$のマルチプレクサ（ＭＵＸ）８０によって受け取られ、これは、クロスバースイッチ３４からのデータ及びコマンド信号発生器８２の出力を選択するように動作する。使用されるネットワークプロトコルに基づき、制御・状態ロジック３６（図２）の指令及び制御のもとで、コマンド記号を周期的に挿入し、送信することが必要となる。ＭＵＸ８０により行われる選択は、出力レジスタ８４へ接続され、そしてそこからポート出力Ｏ（３）を経、そしてポート３が接続されたネットワークリンク１８（図１）を経てＩ／Ｏユニット１４へ送られる。

【００１８】裁定は、各ポート出力において、アービタ８６により行われる。アービタ８６は、各ポート入力Ｉ（０）、Ｉ（１）・・・Ｉ（５）から、対応するポート出力選択ロジック６６（図３）からの要求信号ラインＲ（ｎ）$_3$を受け取る。３つ以上の要求信号が同時にアサートされた場合には、アービタロジック８６は、要求を発しているポート入力の優先順位要求信号をチェックする。全てが非ゼロである場合に、アービタロジック８６は、ラウンドロビン機構の優先順位に基づいて要求を裁定する。しかしながら、競合するポート入力の１つが順番の変更を要求していることがその関連優先順位要求バス（例えば、ポート２の入力ロジック30_2に対するＰＲ（２））上の非ゼロ値によって指示される場合には、アービタロジック８６がそのポート入力へのアクセスを許可する。２つ以上のポート入力がその優先順位要求をアサートする場合には、アービタロジック８６が高レベル優先順位機構に基づいてアクセスを裁定する。２ビットの優先順位要求が等しい場合には、ルート選択が固定優先順位に基づいて行われ、最も高い予め指定された優先順位をもつポート入力にアクセスが許可される。１つの２ビット優先順位バスの値が他のものよりも数値的に大きい場合には、その大きな優先順位要求をアサートするポート入力が次に選択される。

【００１９】裁定が行われると、アービタロジック８６は、６本の信号ライン（各々対応するポート入力の入力ロジック３０に接続される）の１つを経て裁定に勝ったポート入力Ｉ（０）・・・Ｉ（５）へＧＲＡＮＴ信号を発生する。更に、アービタ８６は、クロスバースイッチ３４へ選択信号（ＳＥＬ）を発生し、選択された入力ロジック３０が出力ロジックへルート指定するようにさせ

る。図５は、ポート入力Ｉ（０）に対する入力ロジック３２。のバイアスロジック７０を詳細に示しており、８ビット加算器１４０と８ビット累積レジスタ１４２との組合せが含まれて、実際に、自走モジュロー２５５カウンタを形成する。加算器１４０は、対応するポート入力（ここではポート入力Ｉ（０））に指定されてバイアスレジスタ７２により維持されたバイアス値を受け取り、そのバイアス値を累積レジスタ１４２の内容に加算する。加算器１４０により形成された和は累積レジスタ１４２へ返送され、その内容をＩＦ値だけ増加する。又、累積レジスタの内容は、入力ロジック３０。が参加する各裁定周期ごとにＩＦ値だけ増加される。累積レジスタ１４２の内容が、加算器１４０の巾を越える（即ち、２５５より大きい）点まで増加されると、加算器１４０の桁上げ（Ｃｏ）出力にオーバーフロー信号が発生する。加算器１４０からのオーバーフロー信号は、オーバーフロー検出回路１４６へ送られ、出力（ＯＶ）に応答オーバーフロー信号をアサートし、これは、次いで、信号ライン１４８によって２ビットカウンタ１５０の増加（ＩＮＣ）入力へ接続される。従って、カウンタ１４０の検出されたオーバーフローは、２ビットカウンタ１５０を増加するように働く。２ビットカウンタ１５０の内容は、２ビット優先順位要求値を形成し、これは、２ビット優先順位要求バスＰＲ（０）によって入力ロジック３０。から６つのポート出力Ｏ（０）、Ｏ（１）・・・Ｏ（５）の出力ロジック３２へ搬送される。

【００２０】説明を続ける前に、ポート入力Ｉ（０）に割当てられる入力分数（ＩＦ）は、分数の分子を構成するバイアスレジスタ７２に含まれたバイアス値と、実際には分母である累積レジスタ１４２のサイズとで形成されることに注意するのが有用である。従って、ポート入力Ｉ（０）の場合に、レジスタ７２に保持されたバイアス値が６４（図６について以下に説明する例で使用する値）である場合、ポート入力Ｉ（０）の入力分数は、６４／２５６、即ち１／４となる。２ビットカウンタ１５０は、６入力のオアゲート１５２の出力を受け取る減少（ＤＥＣ）入力を含む。ＧＲＡＮＴ信号は、ポート出力の各々からオアゲート１５２へ送られ、入力としてそこに付与される。関連するポート入力（即ちポート入力Ｉ（０））が、２ビットカウンタ１５０に非ゼロ値を含む状態でポート出力へのアクセスの裁定に参加しそして裁定に勝った場合には、それによりポート出力から生じるＧＲＡＮＴ信号が２ビットカウンタ１５０を減少させる。２ビットカウンタ１５０は、アンダーフローしないように、即ちカウンタの内容がゼロ値であるときに、ＤＥＣ入力がオアゲート１５２の出力を無視しないように設計される。

【００２１】２ビット優先順位要求バスＰＲ（０）は、他のポート入力Ｉ（１）、・・・Ｉ（５）からのバスと共に、６つの優先順位バスＰＲ（ｎ）（ｎ＝０、１、・・・５）を形成し、これらは、ポート入力からの優先順位要求を各ポート出力のアービタロジックユニット８６（図４）へ接続する。又、ここに説明するように、アービタロジックユニットは、３６本の要求ラインＲ（ｎ）。も受け取り、その６つの各々は、６つのポート入力の各々からのものであって、どのポート入力がアクセスに対して張り合っているかをポート出力に識別する要求信号を搬送する。アービタロジック８６は、一般的な従来設計の組合せロジック回路（又はプログラム可能なロジックアレー（ＰＬＡ）エレメント）であり、競合するポートの優先順位バスＰＲ（０）、・・・ＰＲ（５）によって搬送される優先順位要求の状態から、どれがアクセスを受けるべきかを決定し、そしてそのアクセスを、上記のようにクロスバースイッチ３４へ付与されるＳＥＬ信号によりルート指定するように構成される。いずれの２ビットカウンタにもカウントがない場合には、アービタロジックユニットは、ラウンドロビンプロセスに基づいて動作し、そのプロセスによりどれが最後にアクセスを許可されそしてどれが次の順番であるかに基づいて競合ポートの１つを選択する。一方、１つ以上の競合ポートが優先順位要求信号をアサートした場合には、最も高い優先順位要求を有するもの（即ち２ビットカウンタ１５０が最も高いカウントを有するポート入力）にアクセスが許可される。２ビットカウンタ１５０が同じカウントを含んでいる２つ以上のポート入力の優先順位要求間が同点である場合には、アービタは、固定優先順位機構を課し、選択されたポート入力にＧＲＡＮＴ信号を発生する。

【００２２】低レベルラウンドロビン裁定は、通常のメッセージトラフィックに対して使用され、高レベル裁定は、ポートが裁定に参加して不成功であった回数と、指定されたＩＦ値とに基づいて強制的に作用される。高優先順位機構は、ポートの２ビットカウンタ１５０が非ゼロカウントを含むときに入る。図６を参照して本発明の動作を説明する。図６は、裁定周期ＴからＴ＋８及びそれ以上に対しポート０、１及び２（同じポート出力へのアクセスを求める）の裁定を示す。最も左のカラムは、各裁定周期を識別し、他のカラムは、各裁定周期中のレジスタ１４２の内容及び２ビットカウンタの内容（かっこで示す）を表している。裁定周期中に裁定に勝つポートは、ダークの累積値で示されている。ポート０、１及び２の各々に割り当てられたＩＦ値は、各カラムの上部にかっこで示されている。（ここで、出力ポートの帯域巾は、加算器１４０及び累積レジスタ１４２により形成された「カウンタ」のオーバーフロー値で示される。当業者に明らかなように、ポート出力の帯域巾の分割をいかに微細に又は粗くするかと、メッセージトラフィックに対する最大待機とに基づいて、他の値も使用することができる。）

図６は、ポート０、１及び２に対する到来メッセージト

ラフィックのみが特定のポート出力（例えば、ポート出力Ｏ（４））に対して張り合っていると仮定する。明瞭化のため、他のポートは参加しないと仮定し、従って、図示されていない。更に、メッセージトラフィックは各ポートにスタックされ、即ちあるポートに対する到来メッセージトラフィックが裁定されて、再送信のためにポート出力Ｏ（４）にルート指定されたときに、別の到来メッセージが存在するものと仮定する。

【００２３】図６を参照すれば、最初に、第１の裁定周期Ｔの前の時間（Ｔ－ｔ）に、ポート０、１及び２のレジスタ１４２の内容は空である。従って、３つ全ての入力ポート０、１及び２がポート４に向けられたメッセージトラフィックを有すると仮定すれば、ラウンドロビン機構において第１のものである（いずれのカウンタ１５０にもカウントはない）ポート０が裁定周期Ｔの裁定に勝つ。この裁定周期の終わりに、ポート０、１及び２の各レジスタ１４２は、それらの指定バイアス値だけ増加され、従って、次の裁定周期Ｔ＋１については、バイアス値が図示されたようになる。オーバーフローはなくそして関連２ビットカウンタ１５０は空のままであるから、裁定周期Ｔ＋１のラウンドロビン裁定は、ポート４へのアクセスに対しラインの次のポート、即ちポート入力１（図６に太字で示す）を選択する。レジスタ１４２は、再び増加される。ここで、ポート２のレジスタ１４２はオーバーフローを経験してゼロに戻り、それに関連する２ビットカウンタ１５０は、「１」に増加される。従って、次に続く裁定周期Ｔ＋２の間に、アービタロジック８６'（ポート４のポート出力Ｏ（４）の）は、ポート２が２ビットカウンタ１５０にカウントを有するが、他のものは有していないことに注目し、それ故、アービタ周期Ｔ＋２は、ポート２の選択を生じさせる。この裁定周期の終わりに、２ビットカウンタ１５０が１だけ減少され、全てのレジスタ１４２は、指定のバイアス値だけ再び増加される。

【００２４】裁定周期Ｔ＋３は、カウンタ１５０にカウントがないのを見て、再び、ラウンドロビン機構に基づいて、ラインの次のポーチ、即ちポート２にアクセスを許可する。この場合も、レジスタ１４２は、増加される。裁定周期Ｔ＋４は、全てのレジスタ１４２がゼロにロールオーバーしてオーバーフローを生じ、全てのカウンタ１５０が「１」のカウントを含むのを同時に見る。カウンタ１５０のカウントは全て等しく（そして非ゼロであり）、従って、例えば、最初にポート０を見、次いで、ポート１を見、等々といってポート４までラインを下るような固定の優先順位機構に依存することにより、同点状態を突破される。（明らかに、ポート５と他のポートとの間のように、他のポートが常に勝つ。）従って、この場合、ポート０が裁定に勝つ。レジスタ１４２は、対応する指定のバイアス値で再び増加され、一方、ポート０の２ビットカウンタは、当該ポート出力からＤＥＣＲＥＭＥＮＴ信号ラインに信号を発生することにより１だけ減少される。裁定周期Ｔ＋５は、ポート１及び２がそれらのカウンタ１５０に１のカウントを依然含んでいるのを見、他のポートは同点である。この同点状態も、固定の優先順位機構を使用することにより突破され、このときには、ポート２よりも高い固定優先順位を有するポート１が選択され、そしてそのカウンタ１５０が１だけ減少される。

【００２５】裁定周期Ｔ＋６は、ポート２のカウンタが「２」のカウントに増加されるのを見る。これは、そのときカウンタ１５０にカウントを有する唯一のものであるから、ポート４へのアクセスが再び得られ（同時に、低レベルのラウンドロビン機構の次のものであるが）、そしてポート２のカウンタ１５０が減少される。裁定周期Ｔ＋７は、ポート２がそのカウンタ１５０に非ゼロカウントを有している唯一のものとして見、それ故、ポート４へのアクセスが再び選択され、そのカウンタが減少される。この裁定周期の終わりに、ポート０、１、２のレジスタ１４２が増加されると、全てがゼロ値へとロールオーバーしそして全てがそのカウンタ１５０に「１」のカウントを有する。裁定周期Ｔ＋８及びそれに続く周期は、ここで、裁定周期Ｔ＋４、・・・Ｔ＋７を繰り返す。図６を検討することにより明らかなように、このパターンは、ポート２が回数〔１２８／（６４＋６４＋１２８）＝１２８／２５６）の半分だけ裁定に勝つことを示している。一方、ポート０及び１の各々は、回数（６４／２５６）の１／４だけ裁定に勝つ。従って、この機構は、指定バイアス値と２" の比に基づいて帯域巾を割り当てるのに使用され、ここで、ｎは、バイアスされた裁定カウンタ１４２の巾（ここでは、８ビット）である。しかしながら、異なる巾のカウンタを使用して、カウンタ１４２を実施し、帯域巾を割り当てるための比に更に大きな分解能を得ることもできる。更に、６個のポートを有するルータの場合には、２ビットカウンタで充分であるが、６ポートより多くのポートを有するルータは２ビットより大きなカウンタ１５０を必要とする。

【００２６】ラウンドロビン裁定を用いるのではなく、低レベル機構を実施する他の方法もあり、ここに開示した高レベル機構と共に、固定優先順位を使用するか又はメッセージ自体の情報を使用して裁定を行うことができる。更に、高レベル機構において同点状態を打破するために使用される固定優先順位機構は、種々の形態の組合せロジック（例えば、ゲート、プログラム可能なロジックアレー、ルックアップテーブル、等）で実施される他の予め定められた優先順位へ変更することができる。

【図面の簡単な説明】

【図１】一対の中央処理ユニット（ＣＰＵ）を備え、これらは、互いに接続されると共に、本発明によるルート指定装置を用いてメッセージトラフィックを通信するためのシステムエリアネットワーク（ＳＡＮ）により複数

15

の入力／出力（Ｉ／Ｏ）ユニットにも接続されるようなマルチプロセッサシステムの簡単なブロック図である。

【図２】図１のシステムエリアネットワークに使用されるルート指定装置の簡単なブロック図であって、メッセージトラフィックを受け取りそして再送信する多数の個別の入力及びポート出力を有する構造を示す図である。

【図３】図２に示されたルート指定装置の１つのポート入力に関連した入力ロジックを示す簡単なブロック図である。

【図４】図１及び２のルート指定装置の１つのポート出力に関連した出力ロジックを示す簡単なブロック図である。

【図５】待機中メッセージトラフィックを有する図２のポート入力の優先順位を上昇するための優先順位要求を発生するのに使用されるロジックを示すブロック図である。

【図６】メッセージトラフィックを有するポート入力が出力へのアクセスに対して裁定される多数の裁定周期を示す図である。

【符号の説明】

１０　処理システム

16

＊１２ａ、１２ｂ　ＣＰＵ

１４　入力／出力ユニット

１６　ルート指定装置（ルータ）

１８　両方向性通信リンク

３０　入力ロジック

３２　出力ロジック

３４　クロスバースイッチ

３６　状態ロジック

４０　クロックロジック

４２　自己チェック回路

４６　オンラインアクセスポート（ＯＬＡＰ）

５０　入力レジスタ

５２　先入れ先出しバッファ待ち行列（ＦＩＦＯ）

５６　コマンドデコード要素

５８　プロトコル・パケットチェックユニット

６２　弾力性ＦＩＦＯ

６６　出力ポート選択ロジック

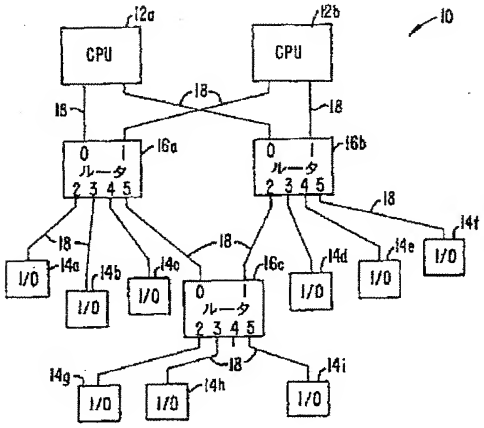
７０　バイアスロジック

８０　マルチプレクサ

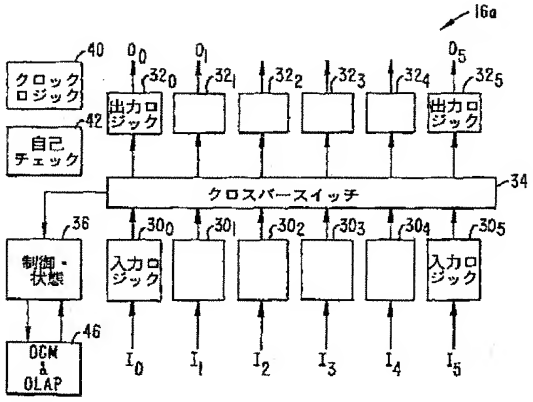
８２　コマンド信号発生器

＊

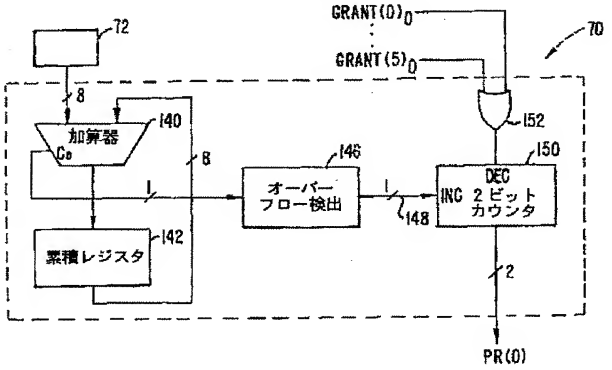
【図１】



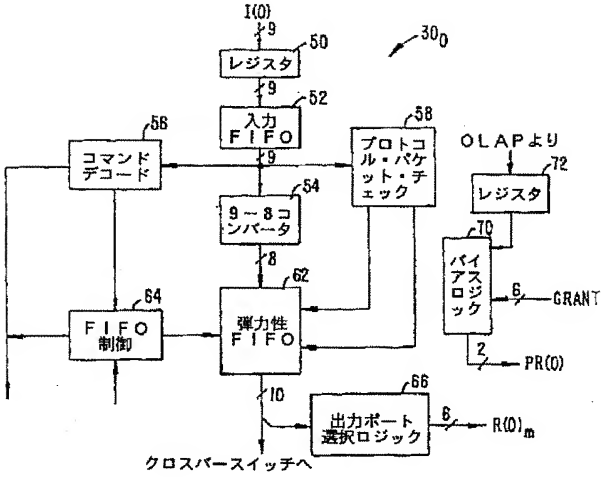
【図２】



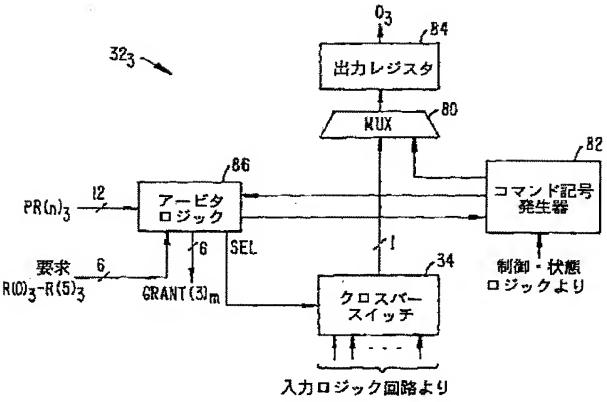
【図５】



【図３】



【図４】



【図６】

裁定周期	(IF=64/256) ポート 入力 I(0)	(IF=64/256) ポート 入力 I(1)	(IF=128/256) ポート 入力 I(2)
T-1	0 [0]	0 [0]	0 [0]
T	0 [0]	0 [0]	0 [0]
T+1	64 [0]	64 [0]	128 [0]
T+2	128 [0]	128 [0]	0 [1]
T+3	192 [0]	192 [0]	128 [0]
T+4	0 [1]	0 [1]	0 [1]
T+5	64 [0]	64 [1]	128 [1]
T+6	128 [0]	128 [0]	0 [2]
T+7	192 [0]	192 [0]	128 [1]
T+8	裁定周期Ｔ＋４からＴ＋７を繰り返す		

フロントページの続き

(72)発明者　ウィリアム　ジョエル　ワトソン
アメリカ合衆国　テキサス州　78756　オースチン　ウルリク　アベニュー　1501

(72)発明者　ディヴィッド　ポール　ソーニア
アメリカ合衆国　テキサス州　78750　オースチン　イメージ　コーヴ　7804